Review of the
Federal Communications Commission's Staff Report of Audit Findings
Audit of the Continuing Property Records of
BellSouth Telecommunications

Ernst & Young LLP has been retained by BellSouth Telecommunications, Inc. to conduct a statistical review of the above-referenced draft audit report on their behalf. Our goal was to determine if there were significant errors and omissions from a statistical and processing standpoint. This work is subject to the Nondisclosure Commitment dated July 27, 1998. The work at Ernst & Young has been directed by Dr. Fritz Scheuren.

Dr. Scheuren, a mathematical statistician and internationally known sampling expert, is the National Technical Director of Statistical Sampling for Ernst & Young LLP. He has extensive sampling background from many years of government service with the IRS and the Social Security Administration and most recently at Ernst & Young. His role has been (and is here) not to take an advocacy position but to make sure that the statistical methods in obtaining and using the data are appropriate for the task at hand.

With this intent in mind, we have analyzed the methods used by the FCC in the continuing property record audit. We find their methodology lacking in a number of respects. The FCC's results, in consequence, do not necessarily present a fair representation of BellSouth's "missing" investment. We summarize our findings below. Technical notes that explain how we arrived at some of our findings are attached as Appendices A, B, and C.

Summary

To give focus to our review comments we will quote the language of the draft report's conclusions (page 14 of Appendix B). Conclusion 2 states:

> "Furthermore, from our sample, we estimate by inference, using a confidence interval of 95 percent, that $291.7 million ± $142.9 million of BST's (BellSouth's) hard-wired COE (central office equipment) cannot be found. Thus, as a result of the audit we expect, with 95 percent certainty, that the costs for hard-wired COE recorded in BST's CPR is overstated by an amount between $148.8 million and $434.6 million."

Based on a number of significant mistakes and other biases, we find that the Commission cannot support this extrapolation. There are at least four problems with their audit approach:

> First, the failure to attempt an investigation of whether equipment was present that was not identifiable in the property records precludes a claim that property is "missing." (The inverse of the audit's investigation of whether there were property records for which the corresponding equipment could not be found.)
>
> Second, the sample design was intended to measure the *proportion of items* not found. It was not intended to measure the *dollar value* of missing items. Had the intent been to make such an inference, the sample should have been designed differently. The audit report's effort to make that jump with the small sample of central office locations actually visited is highly questionable and subject to significant potential sampling error.
>
> Third, the margin of error for missing property investment suggested in the audit report is too narrow. The margin of error was calculated using methods that require a larger sample size than

the one used in the audit. When corrected, the margin of error is so wide that the results provide little useful information concerning the dollar value of plant not found. "Actionable" conclusions concerning statements about dollar amounts of investment should not be made from such results.

Fourth, the methods identified in the audit report contain biases which create inaccuracies and further undermine reliance on the results. These biases include both potential biases in the statistical calculation and in the audit methodology.

Below we discuss each of these errors. Their compound effect significantly undermines the validity of any conclusion that attempts to extrapolate the results of this audit to BellSouth's investment base. Our analysis considered this compounding effect as well as the possible impact attributable to BellSouth's own re-scoring.[1]

Additionally, the staff analysts seem to believe that these errors are justifiable since they claim that their results are corroborated by a Bayesian analysis that was applied to the data. No information regarding important aspects of the Bayesian analysis was given in the staff report. Claims concerning how Bayesian methodology eliminates problems in the initial analysis are made without justification. And final results are given without any discussion of how they were derived. Thus, an unsubstantiated Bayesian analysis, using data containing biases, is being used to corroborate the staff's flawed initial analysis. Because the flaws of the original analysis are implicit in their Bayesian model, their new model provides no additional support.

Lack of a Two-way Audit

The FCC audit only investigated in one direction: sampling from the property records to see if a selected line-item can be located where the property record says it is. If the intent of the audit was to attempt a quantification of "missing" equipment, it would be necessary to conduct a two-way audit. A two-way audit would also include an inventory of randomly selected offices and a check to see if items found in the inventory can be tied back to the property records. The only way to determine "missing" equipment would be to take the results of the initial audit and net them against the results of the reverse direction audit. The failure to conduct the reverse audit here means that the FCC quantification of "missing" investment systematically overstates the actual value and cannot be relied on.

Sample Design (Choosing Line-Items for the Audit)

Related to the issue of the goal of the audit, is the sample design – the plan for choosing the way in which records are selected from the CPR database so that fairly precise estimates can be obtained. By this we mean that the sample should be planned so that resulting confidence intervals will not be too wide with respect to the estimate.

The audit report includes two estimates – one based on the proportion of the number of line items that were or were not found to be in conformance with the rules; the other based on the dollar value of investment not located in the audit. In general, one can get reasonable precision for an estimate of proportion (as the audit report sought in the first estimate) using many different sample designs. Estimates of total dollar values (such as sought in the second estimate) are far more complex and difficult. Results may vary widely under different sample designs, so one should take care when designing for these estimates. Thus, an experienced statistician faced with estimating a proportion and a

[1] A more appropriate methodology would be the use of lower confidence bounds with a high confidence level – 99 percent. The IRS uses this type of lower bound approach in their audit findings. In fact, the IRS calculates estimates in three ways. The method that produces the smallest margin of error is used, and the 95 percent lower confidence bound is the amount assessed. Due to additional non-sampling error and other biases, it is prudent, in our opinion, to use a 99 percent lower confidence bound for the FCC audit.

total dollar amount, would usually plan the sample to obtain a relatively precise estimate for the total dollar amount, and accept the precision that is obtained for the proportion estimate.[2]

The methods identified in the audit report did the exact opposite. The audit sampling plan was designed to produce a precise estimate of the proportion, and accept a relatively imprecise estimate of the total dollar amount. Note, that the margins of error given for the proportion estimates are approximately 3 percent (the plus/minus values given in conclusion 1, page 14 of Appendix B of the report.) However, the reported margins of error for the total investment tied to records of items that were not located are approximately 49 percent of the estimated value. These are very imprecise estimates. Thus, even based on the draft report's own calculations without our corrections (and ignoring the failure to conduct a two-way audit), we find the evidence very weak when making statements about total dollar amounts.

Understated Margin of Error Calculations

The FCC audit report also understates the margin of error of their results. The larger the margin of error, the less reliable the results. The formula employed in the draft report assumes that the sample is large enough to apply a standard (normal) distribution approximation (the "bell-shaped' curve from any introductory statistical text) to obtain confidence intervals.[3] This assumption is wrong. The sample size used in the audit is not large enough to use the normal approximation. This leads the audit report to systematically understate the margin of error in their dollar estimates.

For a simple random sample from the CPR, a sample size of 1,152 (the number of items sampled in the audit) is likely to be large enough for the normal approximation to be appropriate. However, the FCC did not use a simple random sample. Instead central office locations were first randomly selected within groupings (or strata) that the audit staff defined, and then records were randomly selected within chosen locations. In statistical parlance, a two-stage, stratified, cluster design was used.

For this type of design, the total sample size is not as important as the number of locations chosen within each stratum. There were 32 locations. Special advanced techniques need to be used for calculating confidence intervals for these more complex designs.

Based on the FCC's estimates and their standard errors, and correcting for the small sample sizes, we conclude that --

> **The estimate for BellSouth's total investment in error, using FCC scoring with partial credit, is $291.5 million. The one-sided 99 percent lower confidence bound[4] is negative $32.6 million.[5] Notice that the lower confidence bound goes beyond zero. This means that there is no statistically significant difference between the estimated total investment not found and zero.**

These results assume complete acceptance of the audit report's classification of what equipment in the sample should be deemed missing. The only change from the draft report is that we have corrected the margin of error based on simulation results described in Appendix A.

---

[2] One way to have done this is to select many more central office locations, and to stratify the line-items by in-place cost within the selected locations.

[3] For a 95 percent confidence interval this is essentially done by multiplying the standard error (the square root of the variance of the estimator) by 2 and subtracting/adding the result from/to the estimate.

[4] As stated in footnote 1, using a 99 percent confidence bound would more nearly respect the uncertainty of the FCC's results in light of the errors and biases that cannot be quantified.

[5] This uses the published value of the root mean squared error, $72.9 million, and a multiplying factor of 4.447. This factor was found from the results of a simulation described in Appendix A.

The draft report also mentions that the true value of understated investment is "most likely" centered around the "best estimate." Confidence intervals give a measure of the precision of the estimate, but no value contained within the interval is necessarily "better" than any other. In a sense, values within an interval are statistically indistinguishable. Thus, a conservative approach, as would be done in an IRS audit, is to use the lower bound of the interval.[6]

Biased Estimates

Statistical biases exist with regard to the methods used to obtain the estimates and the standard error for the estimates (quantities needed to compute the precision of the estimates). The draft report uses an estimation technique that gives a statistically biased estimate, i.e., the average value of the estimator used by the FCC is not the true population total. The actual bias appears to be negligible, but there is a potential problem with the calculation of the mean squared error – it may understate the true value. However, our simulations suggest that the bias in the mean squared error estimate is also negligible.

An alternative approach which avoids some of these problems is to use an unbiased estimate.[7] This method weights each record in the sample inversely to the probability that the record was selected into the sample. The draft report's biased method does not do this.[8] Biased estimates are appropriate in some circumstances, and it may be appropriate for estimating BellSouth's total investment associated with non-locatable line items. We were concerned, nonetheless, because the number of central office locations sampled is small, so we present estimates based on both methods (see Appendix B).

Aside from statistical calculation issues, there are other sources of potential bias. For example, the audit staff did not use the same team of auditors to inspect each location. When examining the proportion of items found by different audit teams, there are noticeable differences in the scoring of line-items. The audit staff tried to correct this control problem by making "back-at-the-office" changes in the scores. It is unclear whether they succeeded in addressing the original team variability in approach since no locations were revisited to verify that the back office scoring correctly represents the true state of the property records.

In addition to the team and re-scoring effects, we are also concerned about statements in the report that refer to substitutions of sampled items. For example, the FCC draft report states on page 7 of Appendix B that

> "In some instances, the location initially selected was impractical to audit, …. In such cases, another location was randomly selected from that stratum."

If the FCC does not want to audit certain locations, their conclusions should be narrowed accordingly – in fact, just to the records in locations that the FCC was willing to audit. A nice analogy we recently heard is that if you only inventory the shelves in a candy store that you can reach, then you can't make any statements about the candy jars on the shelves that are out of reach.

---

[6] Op. cit., footnote 1.

[7] The technique that we suggest is given in Cochran, W. G. (1967). *Sampling Techniques*, 3rd ed. Wiley, New York, page 303 (equation 11.21) for the estimate; for the variance of the estimate see page 303 (equation 11.24).

[8] Details for the FCC's biased estimate are given in the Cochran reference, page 303 (equation 11.25). We also note that the equation given as the variance of this estimate is really an approximation for the mean square error, page 305, equation 11.30. The mean square error is the appropriate measure of uncertainty to use for a biased estimate. The Cochran formula is only asymptotically correct. The number of locations is too small for its use in this audit situation without some additional checks on its safety.

Bayesian Analysis

The FCC staff performed an additional analysis on the data from the audit sample. The results of this second or "Bayesian" model-based analysis are claimed to be close to the results derived from the earlier "frequentist" or sample design-based analysis. The FCC staff believes that this "closeness" corroborates their findings.

A Bayesian approach relies on a model that, unlike a frequentist approach, employs prior assumptions. Underlying the staff Bayesian arguments is the claim that because the Bayesian approach puts all of their statistical deficiencies into a Bayesian sampling model (see Appendix B), the problems with the audit sample and the derived statistics go away. This, however, is not the case. Like any other model, the outputs are only as valid as the inputs and assumptions. The way in which the staff chooses to use Bayesian methodology camouflages the flaws, but does not remove them.

The following table matches three assumptions that the staff claims are true of a Bayesian sampling methodology to the original flaw that they appear to believe can be overlooked.

No.	Assumption about Bayesian Methodology	Frequentist Problem Staff Believes is Resolved
1	An estimate of the population mean is independent of the choice of sample weights or choice of stratification.	Poor Sample Design, Understated Margin of Error
2	The sample mean is the most likely estimate of the population mean.	Determination of Total Cost in Error to be Removed from the Property Base
3	The method is design free, so the estimator is unbiased.	Biased Results

We will address each of these assumptions in turn. The staff does not provide the details of the Bayesian structure that was used for their analysis, that is, the assumptions and formulas used to calculate their results (see Appendix C for a discussion of what is needed). Thus, our comments on the analysis are based on a common approach of employed when Bayesian methods are employed in survey sampling.

Assumption 1

Bayesian sampling analysis is model oriented. It can, for example, employ a superpopulation from which the finite population – in this case, the CPR database – is a sample. The finite population is physically sampled in order to make inferences.

Several factors make up a Bayesian sampling model, and one of the key factors is to create a probability distribution to incorporate for prior knowledge about characteristics of the population that is being sampled. The staff has not indicated what distribution they are using for their prior knowledge. If it is a model that assumes a lot of prior knowledge, then justification is needed. A model that assumes little to no prior knowledge generally produces results similar to a frequentist analysis (although interpretation of the results may be different).

While some Bayesian sampling models may provide estimates which do not rely on the sample design, they are often not independent of the choice of design elements such as stratification.[9] With different strata, the Bayesian estimates would usually be different too.

[9] See Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin, *Bayesian Data Analysis,* Chapman and Hall, 1998, page 224.

None of this means that the original, frequentist-based analysis was properly done. The sample design is still unsuitable for a precise estimate of the cost associated with unlocatable items, and the margin of error has not been properly calculated. The fact that the staff may have a Bayesian model that provides estimates that are close to improperly calculated values does not provide corroboration.

Assumption 2

In a Bayesian context, the reference to "sample mean" as the most likely value is not clear. We are assuming that the report is referring to the mean of the posterior distribution. While it is true that the Bayesian interpretation of the properties of the estimators is different from a frequentist interpretation, nevertheless, the statement in assumption 2 is not warranted without imposing strong conditions on the prior and the data.

Regardless of what value is "most likely," it is not clear what conclusions, if any, can be drawn from that calculation. In particular, it does not respect the uncertainty in the answer however measured. In our view, the lower bound of a 99% confidence interval remains the best estimate for assessing the total cost of the COE not found.

Assumption 3

The question of statistical bias is technical, and we provide a brief discussion in Appendix C. The elimination of statistical bias in an estimate does not address our main concern of bias induced from nonsampling error. Most of that bias comes from the data collection mechanism. A Bayesian approach is conditioned on the given data collection mechanism; thus, this source of bias still exists and will affect the estimate obtained from the Bayesian model.

Conclusion

The estimates in the draft FCC audit report contain biases and are highly inaccurate in other ways as well. Given these errors and biases, the conclusions in the report concerning the amount of overstated investment are unsound and cannot be fairly relied upon.